

SEQUENCE LISTING

<110> Benning, Christoph

Sanda, Sherrie

Yu, Bin

<120> Compositions and Methods for the Synthesis and Subsequent Modification of Uridine-5-Diphosphosulfoquinovose (UDP-SQ)

<130> MSU-04769

<160> 24

<170> PatentIn version 3.0

<210> 1

<211> 1134

<212> DNA

<213> Cyanobacterium Synechococcus

<220>

<221> CDS

<222> (1) .. (1134)

<400> 1

atg cgc atc gct ctc ttt acc gag acg ttc ctc ccc aaa gtg gat ggc 48
Met Arg Ile Ala Leu Phe Thr Glu Thr Phe Leu Pro Lys Val Asp Gly
1 5 10 15

atc gtc acg cgg ctt cgg cac acg gtc gat cac ctg cag cgt ctt ggc 96
Ile Val Thr Arg Leu Arg His Thr Val Asp His Leu Gln Arg Leu Gly
20 25 30

cac acc gtc atg gtt ttt tgc ccc gac ggc ggg ctc cgc gag cac aag 144
His Thr Val Met Val Phe Cys Pro Asp Gly Gly Leu Arg Glu His Lys
35 40 45

ggg gct cga gtc tat ggg gtt aaa ggc ttt ccg cta ccg ctc tat ccc 192
Gly Ala Arg Val Tyr Gly Val Lys Gly Phe Pro Leu Pro Leu Tyr Pro
50 55 60

gag ctg aag cta gct ttt ccg ttg ccg aaa gtg gga aaa gcc ttg gag 240
Glu Leu Lys Leu Ala Phe Pro Leu Pro Lys Val Gly Lys Ala Leu Glu
65 70 75 80

cgg ttc cgg ccc gac ctg atc cac gtg gtc aat ccg gct gtg ttg ggg 288
Arg Phe Arg Pro Asp Leu Ile His Val Val Asn Pro Ala Val Leu Gly
85 90 95

ttg ggc ggc atc tac tat gcc aag gcg cta aat gtg cca ctc gtg gcg 336
Leu Gly Gly Ile Tyr Tyr Ala Lys Ala Leu Asn Val Pro Leu Val Ala
100 105 110

tcc	tat	cac	acc	cat	ttg	ccg	aaa	tac	ctt	gag	cat	tac	ggg	ctg	ggg	384
Ser	Tyr	His	Thr	His	Leu	Pro	Lys	Tyr	Leu	Glu	His	Tyr	Gly	Leu	Gly	
		115					120					125				
gtc	ttg	gag	ggg	gtg	ctc	tgg	gaa	ttg	ctg	aag	ctg	gcg	cat	aac	caa	432
Val	Leu	Glu	Gly	Val	Leu	Trp	Glu	Leu	Leu	Lys	Leu	Ala	His	Asn	Gln	
	130					135					140					
gca	gcg	atc	aac	ctc	tgt	act	tca	acc	gcg	atg	gtg	cag	gag	ctg	aca	480
Ala	Ala	Ile	Asn	Leu	Cys	Thr	Ser	Thr	Ala	Met	Val	Gln	Glu	Leu	Thr	
	145				150					155						
gat	cac	ggc	att	gag	cac	tgt	tgc	ctc	tgg	cag	cga	gga	gtg	gat	acc	528
Asp	His	Gly	Ile	Glu	His	Cys	Cys	Leu	Trp	Gln	Arg	Gly	Val	Asp	Thr	
				165					170					175		
gag	acc	ttt	cgg	cca	gac	ttg	gct	act	gct	gcg	atg	cgc	gat	cgc	ctc	576
Glu	Thr	Phe	Arg	Pro	Asp	Leu	Ala	Thr	Ala	Ala	Met	Arg	Asp	Arg	Leu	
			180					185					190			
agt	ggc	ggt	aag	ccc	act	gcg	ccc	ttg	ttg	ctc	tac	gtc	gga	cgc	ctc	624
Ser	Gly	Gly	Lys	Pro	Thr	Ala	Pro	Leu	Leu	Leu	Tyr	Val	Gly	Arg	Leu	
		195					200					205				
tca	gcc	gag	aag	caa	atc	gat	cgc	ctg	cga	ccc	att	ttg	gat	gcc	aat	672
Ser	Ala	Glu	Lys	Gln	Ile	Asp	Arg	Leu	Arg	Pro	Ile	Leu	Asp	Ala	Asn	
	210					215					220					
cct	gag	gct	tgc	ttg	gcc	ttg	gtc	ggc	gat	ggc	ccg	cat	cgg	gcc	gaa	720
Pro	Glu	Ala	Cys	Leu	Ala	Leu	Val	Gly	Asp	Gly	Pro	His	Arg	Ala	Glu	
	225				230				235						240	
cta	gag	caa	ttg	ttt	gct	ggc	acc	cag	acg	cag	ttc	att	ggc	tat	ctg	768
Leu	Glu	Gln	Leu	Phe	Ala	Gly	Thr	Gln	Thr	Gln	Phe	Ile	Gly	Tyr	Leu	
				245					250					255		
cat	ggg	gaa	cag	cta	ggg	gcg	gcc	tac	gct	tct	gct	gac	gct	ttt	gtc	816
His	Gly	Glu	Gln	Leu	Gly	Ala	Ala	Tyr	Ala	Ser	Ala	Asp	Ala	Phe	Val	
			260					265					270			
ttt	ccc	tcc	cgg	acc	gaa	acc	ctc	ggt	cta	gtc	ttg	ctg	gaa	gcc	atg	864
Phe	Pro	Ser	Arg	Thr	Glu	Thr	Leu	Gly	Leu	Val	Leu	Leu	Glu	Ala	Met	
		275					280					285				
gca	gcg	ggt	tgt	ccg	gtc	gtg	gcg	gcc	aat	tcc	ggt	ggc	att	ccc	gat	912
Ala	Ala	Gly	Cys	Pro	Val	Val	Ala	Ala	Asn	Ser	Gly	Gly	Ile	Pro	Asp	
	290					295					300					
att	gtc	agc	gac	ggc	att	aat	ggt	ttc	ctg	ttc	gat	cct	gag	gat	gaa	960
Ile	Val	Ser	Asp	Gly	Ile	Asn	Gly	Phe	Leu	Phe	Asp	Pro	Glu	Asp	Glu	
	305				310					315					320	
caa	ggg	gcg	atc	gct	gcg	att	cag	cgc	ttg	ttg	gct	aac	cct	gca	gag	1008
Gln	Gly	Ala	Ile	Ala	Ala	Ile	Gln	Arg	Leu	Leu	Ala	Asn	Pro	Ala	Glu	
				325					330					335		
cgc	gag	att	cta	cgc	caa	gcg	gct	cgt	caa	gaa	gcc	gaa	cgc	tgg	agc	1056
Arg	Glu	Ile	Leu	Arg	Gln	Ala	Ala	Arg	Gln	Glu	Ala	Glu	Arg	Trp	Ser	
			340					345					350			

tgg aac gca gcc acg cgc caa ctc cag gac tac tac tgc gag gtg ttg 1104
 Trp Asn Ala Ala Thr Arg Gln Leu Gln Asp Tyr Tyr Cys Glu Val Leu
 355 360 365

gca gat ggt tgc tta ccc tta gcg gcc tga 1134
 Ala Asp Gly Cys Leu Pro Leu Ala Ala
 370 375

<210> 2

<211> 377

<212> PRT

<213> Cyanobacterium Synechococcus

<400> 2

Met Arg Ile Ala Leu Phe Thr Glu Thr Phe Leu Pro Lys Val Asp Gly
 1 5 10 15

Ile Val Thr Arg Leu Arg His Thr Val Asp His Leu Gln Arg Leu Gly
 20 25 30

His Thr Val Met Val Phe Cys Pro Asp Gly Gly Leu Arg Glu His Lys
 35 40 45

Gly Ala Arg Val Tyr Gly Val Lys Gly Phe Pro Leu Pro Leu Tyr Pro
 50 55 60

Glu Leu Lys Leu Ala Phe Pro Leu Pro Lys Val Gly Lys Ala Leu Glu
 65 70 75 80

Arg Phe Arg Pro Asp Leu Ile His Val Val Asn Pro Ala Val Leu Gly
 85 90 95

Leu Gly Gly Ile Tyr Tyr Ala Lys Ala Leu Asn Val Pro Leu Val Ala
 100 105 110

Ser Tyr His Thr His Leu Pro Lys Tyr Leu Glu His Tyr Gly Leu Gly
 115 120 125

Val Leu Glu Gly Val Leu Trp Glu Leu Leu Lys Leu Ala His Asn Gln
 130 135 140

Ala Ala Ile Asn Leu Cys Thr Ser Thr Ala Met Val Gln Glu Leu Thr
 145 150 155 160

Asp His Gly Ile Glu His Cys Cys Leu Trp Gln Arg Gly Val Asp Thr
 165 170 175

Glu Thr Phe Arg Pro Asp Leu Ala Thr Ala Ala Met Arg Asp Arg Leu
 180 185 190

Ser Gly Gly Lys Pro Thr Ala Pro Leu Leu Leu Tyr Val Gly Arg Leu
 195 200 205

Ser Ala Glu Lys Gln Ile Asp Arg Leu Arg Pro Ile Leu Asp Ala Asn
 210 215 220

Pro Glu Ala Cys Leu Ala Leu Val Gly Asp Gly Pro His Arg Ala Glu
 225 230 235 240

Leu Glu Gln Leu Phe Ala Gly Thr Gln Thr Gln Phe Ile Gly Tyr Leu
 245 250 255

His Gly Glu Gln Leu Gly Ala Ala Tyr Ala Ser Ala Asp Ala Phe Val
 260 265 270

Phe Pro Ser Arg Thr Glu Thr Leu Gly Leu Val Leu Leu Glu Ala Met
 275 280 285

Ala Ala Gly Cys Pro Val Val Ala Ala Asn Ser Gly Gly Ile Pro Asp
 290 295 300

Ile Val Ser Asp Gly Ile Asn Gly Phe Leu Phe Asp Pro Glu Asp Glu
 305 310 315 320

Gln Gly Ala Ile Ala Ala Ile Gln Arg Leu Leu Ala Asn Pro Ala Glu
 325 330 335

Arg Glu Ile Leu Arg Gln Ala Ala Arg Gln Glu Ala Glu Arg Trp Ser
 340 345 350

Trp Asn Ala Ala Thr Arg Gln Leu Gln Asp Tyr Tyr Cys Glu Val Leu
 355 360 365

Ala Asp Gly Cys Leu Pro Leu Ala Ala
 370 375

<210> 3

<211> 2979

<212> DNA

<213> Cyanobacterium Synechococcus

<400> 3
ctacacgtta ccttccggta ctggaaacag tcgtttaatc aaccaattga ttgggtcccaa 60
aacatgaact ttctttttcc tccagaacca aatcgctgca ctgtactgtt cattgcgtat 120
ctttgtcgtt gctgctctcc aatcatatct ctcggtctct tctcttgccg cttttccaat 180
gatctctctt gtttcacggt cgtgcagtaa agttctcagt tttgtcacgc aatcttcaac 240
atctccaggg ttgaacaaaa atccggtttt tccctcctga tgaaaacatc agaatcagaa 300
aaccacaagc tcaatatagg ttgaccata agaacaatca atgcaagatc attttgtgta 360
ccagtctatg attgaataaa gtttcagttc ggttacagct cgcttataag aaaattggca 420
gaaattgttt tttcaaccat ttcggttcgg ttgatatgct catcaatatg gtttggcagt 480
taattgtaat tcagataatt cactgacctg atcttcaggg attatatcag ggattccacc 540
ggcacggggc gcgacgacag gaagtctga agacattgct tcaagaacca caaggccaag 600
tgtctccgac tctgatggca tcacaaacac atctccactt gcgtaagctt gtgagagttc 660
atcgcttgt aacgttccag tgaaaaccgc tggcattccg gtaaacact tctcaagatc 720
ctctctttaa gaaaacgaaa cagataaaca aaattacaat gttgttgact agaaatcttc 780
agataacaat atggccaatc ttttaacaaa ctagtacttg tatgggtccat ctccaatgaa 840
agcaatccga gcttcaggta atttgtccat tacactgcac acaaatttct caatatcaaa 900
attcgataca ccacttaaaa gaagtgaagc cagtttatac aaaattctaa cctctttaaa 960
agctccaaac tcttttctac gccaatgcga cctacatgaa tcactagtgg cttttctggg 1020
tcgccattac tgttaattac aaaatattaa acatcaagat tagcgtggaa agtatcattg 1080
tttttaatgc atataaaaga aacgtatatt ctattcttgc ctcagtctta tacgcatttc 1140
ttgagaacgg aaacggggat tgaagctttc tgaatcgaca cccttattcc aaagtcgaag 1200
ttgattagct tcaatatatg aaaaaaagaa gaagaaaatg taagttttga acaatcatag 1260
agcttgtaaa aaaatgtaat gtatgatctt tcttcacctg cagttgcacc agctgctata 1320
agatcttttc caatggcagc agaaggaact aatgtaagat cagccgctct gtgaaggaac 1380
cctgataaaa gcatattcaa gtttagtttc atattataca tcacaataaa ccagaaaaag 1440
aagaaggaaa ttttgacatt tgaaaagcgg gttttacata cttattatag accacattgg 1500
ttttaccaac caactaaaag tgtatcttgg gatgtatctg cacatcaaca caagcttaat 1560
cttagacaaa aattttttta taacaacatt gtgaaatgag gcagaaaaag gtacttacac 1620
agggacgtgt gtgtggtaag acattactat tggtaacagat agcatttttg ctattgccag 1680
agcaccaaag acctaaaaat tttagtcagg gaaaaaagag agtcaagttc tggattctct 1740
ccagttcact ggtcttgcta atgttttagt aatgtgaatt cttgagggat ttaccataac 1800
tccgggagat gaagcgtgta taatgtcagg cttaaaccgt gcaatttcag agatgattct 1860

tggactaagc	gcaagcgaga	gtggaacctt	ttggtataa	ggacaagggg	agctacagaa	1920
gagaagaaga	aattagcgat	attaccaa	agagaacatc	cagtgagtaa	actaaatgg	1980
gctaccttct	tgatccaatg	actctggctc	cataaaactc	ttcaggaaca	ccttcatgtg	2040
tcgtcacgac	tataacctat	gaagcaaaaa	agttattaaa	aaaaaaaaaa	aaggaacagt	2100
taacacttgt	caagtaattc	taattctgga	aacagttact	tatgagctga	ctgaaaagat	2160
acttaagttg	aagaatgaga	tagtaaaaga	agaaacctcg	tctcccattt	cacggaggt	2220
tctaataaaa	ttctggaatc	tgtttttgta	gccggataca	tagctgcaac	aaaatcaag	2280
agagaatcac	ttccaataat	aacatgacat	ataataaaag	cttttgggtc	atggatcgg	2340
gattccgaga	atcttgggat	attcacaact	aaaatctgac	aactttgact	caaacaaatc	2400
ctgaatgtaa	ttgggtttta	cgatctacta	tataatttgc	taaattgggtg	gtgtagcaaa	2460
ttcatacatt	agcgagtatc	tcttcataaa	ataaatgtaa	cgatcaaata	gaaagaaaaa	2520
aacattacag	gaaaaagtca	accaaggaaa	aaaatgagta	gaatctgttt	tcacagagac	2580
atttcgtcga	acacaaaaca	agcaaaaaaa	gaacactgtg	aagaagactt	acgcaaaggg	2640
agaagggtca	acaaagagag	caattctcct	aggcttagag	agcgactcag	gatcgagaag	2700
cggcgcatca	atctccgatt	cgtcatcttc	tctgacttga	gtaatagtca	tatcgttgga	2760
cccagaaaca	gcttcctttg	taatcactcc	acagaagcgg	agcttgcttt	tcttgctaata	2820
gggtaatctc	cggtgaccaa	acgaaagagg	ggaatgaaga	acaaaagaag	aagacctggg	2880
aggagaacaa	gaggttgcag	aggaagaaga	acaagtgtta	gtcgtgctag	gaagcaaata	2940
aggaggtata	gagagattta	tagaagaaa	agtcgtcat			2979

<210> 4

<211> 2537

<212> DNA

<213> Arabidopsis thaliana

<400> 4

tcataatttg	aaaaagcact	tcttcttctt	cttcttcttc	ttcttcttct	tcttcttccc	60
atcaaaaagt	gctagtggca	agaccgcgca	tattccagta	acaaagttaa	aattcaaaac	120
tatggaacac	aatttgtccc	atgttttggt	gcctgaatta	aacgaccata	tcacttgggt	180
cggccacttc	atctcggata	cgcacaagcg	gttgtcaagg	ttgtaaaaga	cgattttgta	240
aggatctaca	ttggcaaagg	gagctttaca	gacgacttga	aaagtttcgg	tgtgaagatt	300
gaaaggtaaa	acttttggtt	ctttgcactc	ggtgaaccaa	taaagcgacc	catctagata	360
cacaggtgcg	gggtaagcag	caatccgata	aggagcagcg	agagtgcac	acctccaagc	420
attggtacta	aagtcgaaaa	cttcgcatgt	agtagcggtt	tttaggccta	tttctgcaga	480

gttgtataac	caaacgggct	tgtatgtgcc	cctgaatttg	tctttaccga	atccaagcat	540
aaagaatttg	cgtttaagct	tgtagtaacc	atctcgtaag	tcgatcatga	gtttttgata	600
atcgcaaaga	ggaagaggcc	gataccatct	agtggtcggg	ttgaccacat	aaccggattt	660
gtcgtgattg	taaagacaaa	cgagaccgtc	acaactacta	tgtgaaacta	agtacagtac	720
gttatccttc	tccaaggag	tagggatctt	gaacgttgat	gatgaacca	actgcaatgt	780
tcttaaagat	tctatggtcg	ggtttataac	atggtgagga	aatacagaca	ccatcagaac	840
atctggatct	cctgattgct	gacgatgttt	caactaaaat	tgtatacatt	tttattttat	900
taaaaagata	aatatcatth	tgttggaatt	ggaaaacaaa	atctaaaaga	taactttaga	960
tatgtaaggg	tcgcatatgt	atttgattac	attgctatga	gtatatgact	cacttttcca	1020
ataatgaaaa	aaataaagat	gttgagtctt	actaattaag	ggtaaataca	aaattttctga	1080
tcattaatac	aaagaaaaag	cctctcataa	gccaaagcca	ttttactcgc	cgtaaacatc	1140
tccgcagcat	actctttaca	cttccttcct	ctctccgcc	atctctccgc	tccttcagcc	1200
acagcaacct	ccataaccgc	cgtcaaagct	tccacatttg	gcgcaaacat	aaacccaaac	1260
tcatcattca	ccactatagt	cctctttatg	cttgcgtagc	tagatgccat	cactgggtta	1320
ccactcaaca	tcgcttccat	taacgttaaa	tcaagacctt	gtgggtctaag	cgttggattc	1380
acgaacaaat	cgatcccgtt	gtagaaaccc	ttgagctcat	ttgggttttag	agatcccaaa	1440
atggaaactt	tttcccctaa	ttccttgtaa	cgttgctccc	atggccctga	tccagctact	1500
acgaggtaaa	catttgaaata	cgtttggtat	attttcgaga	aagcttcgaa	gagcaatgga	1560
tgtcctttgt	ctttgactaa	tctcccagca	gctcctaaaa	caatcgctga	tgagttttct	1620
ggtaacccta	attttgacct	aaacagagta	cgtagcttct	tgtctgatgt	gaatccgttc	1680
tcgtcgactc	cattgaggat	cacatgaacc	cttttctcag	ggatttggtg	aacgtctctt	1740
agcattttct	cgcagctatc	gctgatagcg	atgtggtgag	cgtagttgtg	gaagaatctg	1800
atttcgtcca	gtatcttggg	aagcacggca	ccgtatagac	ttgcattgaa	gccttgatga	1860
cttggttcgt	ctgggtttacg	gatcaggtct	tggtaaatac	ttgattgtaa	gctctctaac	1920
gcaatgccgt	gccaggatac	agcgagggtt	ggaacctccc	ggcgatcca	gtgaggtaaa	1980
gcaacacttt	cagagtgaac	cgcacgaaa	ggttctttct	tgttttcttc	ttggtaaagc	2040
tcccatgcct	tgttgtaccg	ccattttccg	ggctccgcgt	cgccgtgaga	atggattata	2100
gggtaaataga	tttggtcggg	aacagggggg	attttgttgg	tttcagggga	ttgggtctaag	2160
ggagaggtga	aaacgtggac	acggtgtcca	cggcgagcta	aggcgggtga	tagagtgaag	2220
gcgtggcggt	ccatgccgcc	aggattggga	cccgtgggcc	atttcgggga	gaaaacagcg	2280
agtttttagtg	ttttgggagg	agggttggtt	aaggagaaat	caaggcggtt	ccaggcaaat	2340

tgggcggttt	ggaggtcgcc	agaccacggt	ggctggtttg	tgtcagagga	ggaggaggag	2400
actgcagcgg	tggaggagga	gcaagtggag	gtgcggagga	gaaagagagc	cggatatggta	2460
aagaggacgg	tgaagaagag	gaaagtacaa	aggctaaaat	gagaattagg	tttcttgagt	2520
ttggttttgtg	aagccat					2537

<210> 5

<211> 2369

<212> DNA

<213> *Arabidopsis thaliana*

<400> 5

ttaaggtctc	atgcatttga	ccagaacatc	gacgaatcgc	ttgtacatat	gttgcttcat	60
gtacattttc	tcaaccattt	tgcgtccttc	actccccaat	cgtagcctct	cgtctggatt	120
cctaagtaga	tacaagagat	tatgagctaa	ttctttgtta	cggatctcc	ccattgagtg	180
aagtagacca	gtcatgttgt	gttgaaccat	ctctttgggt	cctcctgcat	ctgttccac	240
cactgcaagt	ccataagcca	ttgcttcgat	tgtcactcta	ccaaatgttt	caccaactcc	300
ctgtaaaaaa	tcaaggttca	aaataaataa	accgaacaaa	atcgaaatgg	tcaaaccgaa	360
atcaaaacct	acattttgaa	caaaaacaaa	actgaaatgt	tttatccaaa	accaaactga	420
aatcaatcac	aaattgggtt	atttgattca	gtaattcaag	ttcctataaa	actgataaaa	480
ctaaaccaa	acctaactgt	aaactacttt	ctagtggaaa	actgcatatg	catcaaataa	540
tgtttttgag	gtgtgaggag	gattacctgg	gagtttgtaa	cgtagacatc	tgctgcagag	600
tataatgaag	caacacgggt	tgttgcagga	gtccacatta	ccgacttaga	taagtttccg	660
ctgtttgaca	agaagcttaa	catctcttta	acgtatccaa	ctttgttgct	ctttgaaccc	720
acggatccta	aaagaacttt	aagttcttgc	ttctctcttc	ttaagccatt	gtcaagtgtg	780
agagaaacac	ttttcatctg	tcgtgatgaa	cctcttaaac	gatgtttgct	ggagagacta	840
accttttctt	tcctaattgat	ccccttgatg	ttcctttgag	attcttgctc	tctctcagaa	900
agagccaagg	caatagattc	aaggagaaga	agttgtccct	ttggtgggtt	tatgctgcta	960
agagacatca	caagcatatc	tgaatctgtt	attcctaact	ctgttctcac	tgattcgcgt	1020
aatatttgtc	tcttcaccct	cattttctct	ggtgaaagtg	ttggtgtgtt	gagtgaagaa	1080
ggaatccccg	ctacaaaagc	taactcatca	ttaacagata	gtggaacaat	cactggttgt	1140
gatctaagct	ttatatgctc	ctcctcgcac	catgttaacc	attgtctgct	ctgtgattca	1200
gataagaaaa	tcagcatttt	cactcgggtc	agaactgggt	tcgctcgatc	aaagtattct	1260
cgtcgattct	ccattatcca	ccaagctatt	tgacttccac	cagctggatg	atgattcatg	1320
tattgatctg	cgagaaaagg	aaaaaaaaaa	agtaattaag	atatgctttc	ctctgattgt	1380


```

aaattagtag tctaaaaact gcataatgaa ttcttaccta tccatgaggt acacactgct 1440
gatcctgcga tgatcaaatac tgctttcatg gcagtcttga agctgagttc tcctttgtct 1500
tcaacaactt tgatccttct cctactgagc tcttgcatca atccacctct cctgctaaga 1560
actactgcag agactggtgc accacagctc aaaagctctg aagccagctc catcatagaa 1620
attggagcac cagtcattga tagctcgtgg aaaagcagga cgaatctcct tgaccaaaca 1680
aggcgtttaa aatctgattt cctgtcgcaa gtcccagatc ttctatgcgg gctccattca 1740
agaactttat cctctagtga tccaaaggga ccaagaagct taccataagt agcatttgct 1800
aatggaagtt gtggatcttg ctcatcatcc aaatccttag tctcaagtac ctctttaatc 1860
accttctgct taacacggat cttactacga gaagtccgga cagtttttct cgtcttctgc 1920
ttggaactca agctccgtcg agaaacccca tcatctttct tgatcagact aacatccgtc 1980
ctcttattcg aaccagcatc atctttacca gtaatattaa ccaaagcctc agaatttctca 2040
ttagcaacaa catccagtcc tttaatcttc tccatatata actcatcctt tctcgggtctg 2100
cctccaaacc gtaaaaaactc aactttgctt tcattatcat gtgccacct agactgaaca 2160
taaaatccaa gatacgcca aagcgtaatc aaaagcagcc aataaaccaa ccggctacta 2220
cgaaaccact gaaacgctcc tcctccacca tgacctctac gcggagtcct accagaatac 2280
actctagggtg tacccttgg agtagacctc cctgacagtg aagacttaac acttgtctgt 2340
ctcagcgggc ataaccgaat ctctccat 2369

```

<210> 6

<211> 1799

<212> DNA

<213> Arabidopsis thaliana

<220>

<221> CDS

<222> (170)..(1603)

<400> 6

```

gtcgaccac gcgtccgctc atctctcatc gttccgggag aagagaagag agaccatcc 60
ctcacttcaa agttcaaagt ctcgaaggat cttctccaac tctctctaaa caagattcca 120
aattttcaaa ggtgaatttg tttgatagaa tcaagaacaa acctttaaa atg gcg cat 178
                                     Met Ala His
                                     1
cta ctt tca gct tca tgc cct tca gtt atc tca ctt agc agc agc agc 226
Leu Leu Ser Ala Ser Cys Pro Ser Val Ile Ser Leu Ser Ser Ser Ser
      5              10              15

```

agc aag aat tca gtt aag ccg ttt gtt tca ggg cag acc ttc ttc aat	274
Ser Lys Asn Ser Val Lys Pro Phe Val Ser Gly Gln Thr Phe Phe Asn	
20 25 30 35	
gct cag ctt ctt tca aga tct tct ctc aaa gga ctt ctc ttc caa gag	322
Ala Gln Leu Leu Ser Arg Ser Ser Leu Lys Gly Leu Leu Phe Gln Glu	
40 45 50	
aag aaa ccg aga aaa agc tgc gtt ttc aga gca act gct gta cct ata	370
Lys Lys Pro Arg Lys Ser Cys Val Phe Arg Ala Thr Ala Val Pro Ile	
55 60 65	
acc caa caa gca cca ccc gaa aca tct acc aat aac tca tcc tct aaa	418
Thr Gln Gln Ala Pro Pro Glu Thr Ser Thr Asn Asn Ser Ser Ser Lys	
70 75 80	
cca aag cgt gtt atg gtc att ggt gga gat ggt tat tgc ggt tgg gct	466
Pro Lys Arg Val Met Val Ile Gly Gly Asp Gly Tyr Cys Gly Trp Ala	
85 90 95	
act gct ctc cac ttg tcc aag aag aat tac gaa gtt tgc att gtt gac	514
Thr Ala Leu His Leu Ser Lys Lys Asn Tyr Glu Val Cys Ile Val Asp	
100 105 110 115	
aac ctt gta aga cgt ctt ttc gac cac cag ctt gga ctt gag tca ttg	562
Asn Leu Val Arg Arg Leu Phe Asp His Gln Leu Gly Leu Glu Ser Leu	
120 125 130	
act cct att gcc tcc att cat gac cga atc agc cga tgg aag gct ttg	610
Thr Pro Ile Ala Ser Ile His Asp Arg Ile Ser Arg Trp Lys Ala Leu	
135 140 145	
aca ggg aaa tca att gag ttg tac gtt ggt gat atc tgt gat ttc gaa	658
Thr Gly Lys Ser Ile Glu Leu Tyr Val Gly Asp Ile Cys Asp Phe Glu	
150 155 160	
ttc tta gct gag tct ttc aag tct ttt gag ccg gat tca gtt gtc cac	706
Phe Leu Ala Glu Ser Phe Lys Ser Phe Glu Pro Asp Ser Val Val His	
165 170 175	
ttt ggg gaa cag aga tcc gct cct tac tcg atg att gac cgg tcc aga	754
Phe Gly Glu Gln Arg Ser Ala Pro Tyr Ser Met Ile Asp Arg Ser Arg	
180 185 190 195	
gca gtt tat aca cag cac aac aat gtg att ggg act ctc aac gtt ctc	802
Ala Val Tyr Thr Gln His Asn Asn Val Ile Gly Thr Leu Asn Val Leu	
200 205 210	
ttt gct ata aaa gag ttt gga gag gag tgt cat ctt gta aaa ctt ggg	850
Phe Ala Ile Lys Glu Phe Gly Glu Glu Cys His Leu Val Lys Leu Gly	
215 220 225	
acg atg ggt gag tat gga act cca aat att gac atc gag gaa ggt tat	898
Thr Met Gly Glu Tyr Gly Thr Pro Asn Ile Asp Ile Glu Glu Gly Tyr	
230 235 240	

ata acc ata acc cac aac ggt aga act gac act ttg cca tac ccc aag Ile Thr Ile Thr His Asn Gly Arg Thr Asp Thr Leu Pro Tyr Pro Lys 245 250 255	946
caa gct agc tcc ttt tat cat ctt agc aaa gtt cat gat tgc cac aac Gln Ala Ser Ser Phe Tyr His Leu Ser Lys Val His Asp Ser His Asn 260 265 270 275	994
att gct ttt act tgc aag gct tgg ggt att aga gcc act gat ctc aac Ile Ala Phe Thr Cys Lys Ala Trp Gly Ile Arg Ala Thr Asp Leu Asn 280 285 290	1042
caa gga gtt gtt tat gga gtg aag act gat gag aca gag atg cat gag Gln Gly Val Val Tyr Gly Val Lys Thr Asp Glu Thr Glu Met His Glu 295 300 305	1090
gaa ctc cgt aac cga ctg gat tac gat gct gtg ttt ggt aca gca ctt Glu Leu Arg Asn Arg Leu Asp Tyr Asp Ala Val Phe Gly Thr Ala Leu 310 315 320	1138
aac cgg ttc tgt gtg caa gct gct gtt ggt cac cca ctt aca gtt tat Asn Arg Phe Cys Val Gln Ala Ala Val Gly His Pro Leu Thr Val Tyr 325 330 335	1186
ggt aaa ggt ggt cag acg aga ggc tac ctc gat ata aga gac acg gtt Gly Lys Gly Gly Gln Thr Arg Gly Tyr Leu Asp Ile Arg Asp Thr Val 340 345 350 355	1234
caa tgt gtt gag atc gct ata gca aac ccg gca aaa gct ggt gag ttc Gln Cys Val Glu Ile Ala Ile Ala Asn Pro Ala Lys Ala Gly Glu Phe 360 365 370	1282
cgg gtc ttc aac caa ttt aca gaa cag ttt tca gtc aat gaa ctg gct Arg Val Phe Asn Gln Phe Thr Glu Gln Phe Ser Val Asn Glu Leu Ala 375 380 385	1330
tca ctc gtc act aaa gcg ggt tca aag ctt ggg cta gac gtg aaa aag Ser Leu Val Thr Lys Ala Gly Ser Lys Leu Gly Leu Asp Val Lys Lys 390 395 400	1378
atg acg gtg cct aac ccg aga gtg gag gca gaa gaa cat tac tac aac Met Thr Val Pro Asn Pro Arg Val Glu Ala Glu Glu His Tyr Tyr Asn 405 410 415	1426
gca aag cac act aag ctg atg gaa ctt gga ctt gag cct cac tat cta Ala Lys His Thr Lys Leu Met Glu Leu Gly Leu Glu Pro His Tyr Leu 420 425 430 435	1474
tct gac tca ctt ctt gat tgc ttg ctc aac ttt gct gtt cag ttt aaa Ser Asp Ser Leu Leu Asp Ser Leu Leu Asn Phe Ala Val Gln Phe Lys 440 445 450	1522
gat cgt gtg gac acg aaa caa atc atg cct agt gtt tcc tgg aag aag Asp Arg Val Asp Thr Lys Gln Ile Met Pro Ser Val Ser Trp Lys Lys 455 460 465	1570
att ggc gtc aag act aag tcc atg acc aca taa agtgcagacc aatattacac Ile Gly Val Lys Thr Lys Ser Met Thr Thr 470 475	1623
ataaggagag attatgaaag agatgatgtg ttgttttgta tcttcaaact tcatttctgc	1683

aaaagacttg ctaggcttaa gaggttttgt ccatattaca ttgtgcaggt tctttaatgt 1743

tagatcttaa tttcgatgaa aaaaaaaaaa aaaaaaaaaa aaaaaagggc ggccgc 1799

<210> 7

<211> 477

<212> PRT

<213> Arabidopsis thaliana

<400> 7

Met Ala His Leu Leu Ser Ala Ser Cys Pro Ser Val Ile Ser Leu Ser
1 5 10 15

Ser Ser Ser Ser Lys Asn Ser Val Lys Pro Phe Val Ser Gly Gln Thr
20 25 30

Phe Phe Asn Ala Gln Leu Leu Ser Arg Ser Ser Leu Lys Gly Leu Leu
35 40 45

Phe Gln Glu Lys Lys Pro Arg Lys Ser Cys Val Phe Arg Ala Thr Ala
50 55 60

Val Pro Ile Thr Gln Gln Ala Pro Pro Glu Thr Ser Thr Asn Asn Ser
65 70 75 80

Ser Ser Lys Pro Lys Arg Val Met Val Ile Gly Gly Asp Gly Tyr Cys
85 90 95

Gly Trp Ala Thr Ala Leu His Leu Ser Lys Lys Asn Tyr Glu Val Cys
100 105 110

Ile Val Asp Asn Leu Val Arg Arg Leu Phe Asp His Gln Leu Gly Leu
115 120 125

Glu Ser Leu Thr Pro Ile Ala Ser Ile His Asp Arg Ile Ser Arg Trp
130 135 140

Lys Ala Leu Thr Gly Lys Ser Ile Glu Leu Tyr Val Gly Asp Ile Cys
145 150 155 160

Asp Phe Glu Phe Leu Ala Glu Ser Phe Lys Ser Phe Glu Pro Asp Ser
165 170 175

Val Val His Phe Gly Glu Gln Arg Ser Ala Pro Tyr Ser Met Ile Asp
180 185 190

Arg Ser Arg Ala Val Tyr Thr Gln His Asn Asn Val Ile Gly Thr Leu
 195 200 205
 Asn Val Leu Phe Ala Ile Lys Glu Phe Gly Glu Glu Cys His Leu Val
 210 215 220
 Lys Leu Gly Thr Met Gly Glu Tyr Gly Thr Pro Asn Ile Asp Ile Glu
 225 230 235 240
 Glu Gly Tyr Ile Thr Ile Thr His Asn Gly Arg Thr Asp Thr Leu Pro
 245 250 255
 Tyr Pro Lys Gln Ala Ser Ser Phe Tyr His Leu Ser Lys Val His Asp
 260 265 270
 Ser His Asn Ile Ala Phe Thr Cys Lys Ala Trp Gly Ile Arg Ala Thr
 275 280 285
 Asp Leu Asn Gln Gly Val Val Tyr Gly Val Lys Thr Asp Glu Thr Glu
 290 295 300
 Met His Glu Glu Leu Arg Asn Arg Leu Asp Tyr Asp Ala Val Phe Gly
 305 310 315 320
 Thr Ala Leu Asn Arg Phe Cys Val Gln Ala Ala Val Gly His Pro Leu
 325 330 335
 Thr Val Tyr Gly Lys Gly Gly Gln Thr Arg Gly Tyr Leu Asp Ile Arg
 340 345 350
 Asp Thr Val Gln Cys Val Glu Ile Ala Ile Ala Asn Pro Ala Lys Ala
 355 360 365
 Gly Glu Phe Arg Val Phe Asn Gln Phe Thr Glu Gln Phe Ser Val Asn
 370 375 380
 Glu Leu Ala Ser Leu Val Thr Lys Ala Gly Ser Lys Leu Gly Leu Asp
 385 390 395 400
 Val Lys Lys Met Thr Val Pro Asn Pro Arg Val Glu Ala Glu Glu His
 405 410 415
 Tyr Tyr Asn Ala Lys His Thr Lys Leu Met Glu Leu Gly Leu Glu Pro
 420 425 430
 His Tyr Leu Ser Asp Ser Leu Leu Asp Ser Leu Leu Asn Phe Ala Val
 435 440 445

Gln Phe Lys Asp Arg Val Asp Thr Lys Gln Ile Met Pro Ser Val Ser
 450 455 460

Trp Lys Lys Ile Gly Val Lys Thr Lys Ser Met Thr Thr
 465 470 475

<210> 8

<211> 1209

<212> DNA

<213> Cyanobacterium Synechococcus

<400> 8

```

gtgaagattc ttgtattggg tggcgatggg ttctgcgggt ggccctgcgc tctcaatttg      60
gctgctgcag gtcacgccgt caccattgtt gacaacctcg ttcgccgcaa gacagacgtg      120
gaattggggg ttcagtcctt cactccgatc gcgacgattg aacgccgggt gaaggcatgg      180
caagaaacgg gcgggcagcc gattagcttt gtcaatctcg acttagcggc tgattacgat      240
cgcctctgtg cactactgct agaaacgcag ccggatgcga tcgtgcattt tgccgaacag      300
cgcgccgccc cctattcaat gaagagtgca tggcataagc gcttcacggg caataacaac      360
gtcaacgcca cccacaatct gctctgcgcc tgtgtggatg tcggcctcaa gtcccacatt      420
gtccacttgg gcaccatggg cgtctatgga tacggtagcc atcgcggggc tacgattcct      480
gaaggctact tagaagtgga agtcgtccag cgggatggcc aacgctttga agagaagatt      540
cttcaccccg ttgatccggg tagcgtctat cacatgacca agacgctgga tcaattgttg      600
ttctactact acaacaagaa cgacaacatc caagtcaccg accttcacca gggattgtgc      660
tggggcacga acaccgatca ctgtaatctc caccggatc tgaccaatcg gttcgactac      720
gacggtgatt acggcacagt cttgaaccgc ttcttgatgc aggcggcgat cggctatccc      780
ttgactgtgc atggcgttgg tggccaaacc cgagccttca tccacattcg cgactcagtg      840
cgctgcgtcc aactggcgat cgaaaatccg ccagcagcca atgaaaaagt ccgcatcttt      900
aaccagatga cggaaaccta ccaagtcaag gatttggcag agaaagtggc ggcattgacc      960
ggtgctgaaa tcgcctacct gcccaatcca cgcaaggaag cccttgagaa cgacttgatt     1020
gtcgacaacc gctgcttgat tgatttaggc ctcaatccga ccaccttgga caatggcctg     1080
atgagcgaag tggtagaaat tgcgcagaag ttgccgatc gctgcgatcg cgccaaaatt     1140
ccctgcgttt ctgcctggac ccgtaatcaa gctgaagctc tcagcgctcc tgaaaccgct     1200
ctgcgctaa                                     1209

```

<210> 9

<211> 402

<212> PRT

<213> Cyanobacterium Synechococcus

<400> 9

Met	Lys	Ile	Leu	Val	Leu	Gly	Gly	Asp	Gly	Phe	Cys	Gly	Trp	Pro	Cys		
1				5					10					15			
Ala	Leu	Asn	Leu	Ala	Ala	Ala	Gly	His	Ala	Val	Thr	Ile	Val	Asp	Asn		
			20					25					30				
Leu	Val	Arg	Arg	Lys	Thr	Asp	Val	Glu	Leu	Gly	Val	Gln	Ser	Leu	Thr		
		35					40					45					
Pro	Ile	Ala	Thr	Ile	Glu	Arg	Arg	Leu	Lys	Ala	Trp	Gln	Glu	Thr	Gly		
	50					55					60						
Gly	Gln	Pro	Ile	Ser	Phe	Val	Asn	Leu	Asp	Leu	Ala	Ala	Asp	Tyr	Asp		
65					70					75					80		
Arg	Leu	Cys	Ala	Leu	Leu	Leu	Glu	Thr	Gln	Pro	Asp	Ala	Ile	Val	His		
				85					90					95			
Phe	Ala	Glu	Gln	Arg	Ala	Ala	Pro	Tyr	Ser	Met	Lys	Ser	Ala	Trp	His		
			100					105					110				
Lys	Arg	Phe	Thr	Val	Asn	Asn	Asn	Val	Asn	Ala	Thr	His	Asn	Leu	Leu		
		115					120					125					
Cys	Ala	Cys	Val	Asp	Val	Gly	Leu	Lys	Ser	His	Ile	Val	His	Leu	Gly		
	130					135					140						
Thr	Met	Gly	Val	Tyr	Gly	Tyr	Gly	Ser	His	Arg	Gly	Ala	Thr	Ile	Pro		
145					150				155						160		
Glu	Gly	Tyr	Leu	Glu	Val	Glu	Val	Val	Gln	Arg	Asp	Gly	Gln	Arg	Phe		
				165					170					175			
Glu	Glu	Lys	Ile	Leu	His	Pro	Val	Asp	Pro	Gly	Ser	Val	Tyr	His	Met		
			180					185					190				
Thr	Lys	Thr	Leu	Asp	Gln	Leu	Leu	Phe	Tyr	Tyr	Tyr	Asn	Lys	Asn	Asp		
		195					200					205					
Asn	Ile	Gln	Val	Thr	Asp	Leu	His	Gln	Gly	Ile	Val	Trp	Gly	Thr	Asn		
	210					215					220						
Thr	Asp	His	Cys	Asn	Leu	His	Pro	Asp	Leu	Thr	Asn	Arg	Phe	Asp	Tyr		
225					230					235					240		
Asp	Gly	Asp	Tyr	Gly	Thr	Val	Leu	Asn	Arg	Phe	Leu	Met	Gln	Ala	Ala		
				245					250					255			

Ile Gly Tyr Pro Leu Thr Val His Gly Val Gly Gly Gln Thr Arg Ala
 260 265 270
 Phe Ile His Ile Arg Asp Ser Val Arg Cys Val Gln Leu Ala Ile Glu
 275 280 285
 Asn Pro Pro Ala Ala Asn Glu Lys Val Arg Ile Phe Asn Gln Met Thr
 290 295 300
 Glu Thr Tyr Gln Val Lys Asp Leu Ala Glu Lys Val Ala Ala Leu Thr
 305 310 315 320
 Gly Ala Glu Ile Ala Tyr Leu Pro Asn Pro Arg Lys Glu Ala Leu Glu
 325 330 335
 Asn Asp Leu Ile Val Asp Asn Arg Cys Leu Ile Asp Leu Gly Leu Asn
 340 345 350
 Pro Thr Thr Leu Asp Asn Gly Leu Met Ser Glu Val Val Glu Ile Ala
 355 360 365
 Gln Lys Phe Ala Asp Arg Cys Asp Arg Ala Lys Ile Pro Cys Val Ser
 370 375 380
 Ala Trp Thr Arg Asn Gln Ala Glu Ala Leu Ser Ala Pro Glu Thr Ala
 385 390 395 400

Leu Arg

<210> 10

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<220>

<221> misc_feature

<223> Synthetic

<400> 10

aaaggatccc gtgttatggt cattgg

26

<210> 11

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<220>
 <221> misc_feature
 <223> Synthetic
 <400> 11
 gtcggatcct tatgtggtca tggact 26
 <210> 12
 <211> 24
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Synthetic
 <220>
 <221> misc_feature
 <223> Synthetic
 <400> 12
 tttggatccc gcatcgctct cttt 24
 <210> 13
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Synthetic
 <220>
 <221> misc_feature
 <223> Synthetic
 <400> 13
 ataagcttcg agctcaggcc gct 23
 <210> 14
 <211> 29
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Synthetic

<220>
 <221> misc_feature
 <223> Synthetic
 <400> 14
 cgggatccat gacgactctt tcttctata 29
 <210> 15
 <211> 28
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Synthetic
 <220>
 <221> misc_feature
 <223> Synthetic
 <400> 15
 aaggatccct acacgttacc ttccgga 28
 <210> 16
 <211> 28
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Synthetic
 <220>
 <221> misc_feature
 <223> Synthetic
 <400> 16
 aaggatccat ggcttcacaa accaaact 28
 <210> 17
 <211> 28
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Synthetic

<220>
 <221> misc_feature
 <223> Synthetic
 <400> 17
 gcggatcctc atattttgaa aaagcact 28
 <210> 18
 <211> 27
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Synthetic
 <220>
 <221> misc_feature
 <223> Synthetic
 <400> 18
 agggtagcat ggagggattc gggtatc 27
 <210> 19
 <211> 28
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Synthetic
 <220>
 <221> misc_feature
 <223> Synthetic
 <400> 19
 gcggtagcctt aaggtctatg catttgac 28
 <210> 20
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Synthetic

<220>
 <221> misc_feature
 <223> Synthetic
 <400> 20
 ctaggtacca aatggcgcat ctactt 26
 <210> 21
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Synthetic
 <220>
 <221> misc_feature
 <223> Synthetic
 <400> 21
 aaggatcctg cgctaaagtc gcactc 26
 <210> 22
 <211> 71
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Synthetic
 <220>
 <221> misc_feature
 <222> (12)..(12)
 <223> The residue at this position is linked to six contiguous
 histidine residues
 <400> 22
 atgagaggat cgggatccgc atgcgagctc ggtaccccg gtcgacctgc agccaagctt 60
 aattagctga g 71

<210> 23
 <211> 73
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Synthetic
 <220>
 <221> misc_feature
 <222> (12)..(12)
 <223> The residue at this position is linked to six contiguous
 histidine residues
 <400> 23
 atgagaggat ctacggatcc gcatgcgagc tcggtacccc ggtcgacct gcagccaagc 60
 ttaattagct gag 73
 <210> 24
 <211> 72
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Synthetic
 <220>
 <221> misc_feature
 <222> (12)..(12)
 <223> The residue at this position is linked to six contiguous
 histidine residues
 <400> 24
 atgagaggat ctgggatccg catgcgagct cggtaccccc ggtcgacctg cagccaagct 60
 taattagctg ag 72

26
 ant